Seeing is believing: Structure of the catalytic domain of HIV-1 integrase in complex with human LEDGF/p75

Christina Marchetti Bradley and Robert Craigie*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 5 Center Drive, Bethesda, MD 20892

Retroviruses covalently insert their genome into the DNA of the host cell and subsequently coopt cellular machinery for DNA replication, transcription, and protein expression (1). These viruses also exploit cellular proteins to assist in this stable insertion of their genetic material into the host genome, a process called integration. In a recent issue of PNAS, Cherepanov *et al.* (2) reported the first structure of a retroviral integrase, the viral enzyme that catalyzes integration, in complex with a host protein. This definitive structural evidence for an interaction between HIV-1 integrase and any other protein adds to the significant evidence for the role of host proteins in integration that has been accumulated over the past decade (ref. 3 and references therein).

The infectious particles of retroviruses, called virions, contain two RNA copies of their genomes. After viral entry and a series of poorly understood uncoating steps, the RNA is released into the host cytoplasm. Here the viral enzyme reverse transcriptase synthesizes a double-stranded DNA copy of the genome by using the RNA as a template. The next step is the hallmark of the retroviral life cycle: the viral cDNA is transported to the nucleus and is inserted into the host genome (1). Integration is required for infection and ensures the stable association of the viral genome in the host cell for subsequent generations. From its new position, the viral genome is transcribed, leading to the synthesis of viral proteins and full-length transcripts of the genome and, ultimately, to new virus particles.

The chemistry of integration is catalyzed by the viral enzyme integrase, many copies of which are found in the virion (1). The integration reaction has been successfully recapitulated *in vitro* with recombinant integrase and short DNA oligonucleotides representing the viral DNA ends and the target DNA. With this approach, the biochemical mechanism of integration has been elucidated. First, a pair of dinucleotides at the 3′ ends of the viral DNA are cleaved, exposing the conserved CA sequence that marks the boundary between the host DNA and the integrated viral cDNA. Next, during strand transfer, the 3′ hydroxyl groups are joined to opposite strands of the host DNA at sites separated by 5 bp in the case of HIV-1. Cellular enzymes repair the resulting intermediate to complete integration.

Integration takes place in a more complex environment than these biochemical assays suggest. Before integration, the viral cDNA is associated with a number of viral and cellular proteins in a large nucleoprotein assembly called the preintegration complex (PIC). Many of the protein components of the PIC are derived from the core of the infecting virion, but some are acquired from the cytoplasm of the infected cell. The viral proteins associated with the HIV-1 PICs include integrase, reverse transcriptase, matrix, and Vpr (4–6). In addition, the list of host proteins that have been reported to associate with HIV-1 PICs or to interact directly with HIV-1 integrase continues to grow. These proteins include INI1, HMGA1, BAF, EED, p300, and LEDGF/p75, the topic of this Commentary (reviewed in ref. 3). The role of most of these cellular proteins in integration is largely unknown.

This definitive structural study adds to the significant evidence for the role of host proteins in integration.

Expression of HIV-1 integrase in human cells from a synthetic gene led to the discovery that integrase stably associates with LEDGF/p75 (lens epithelium-derived growth factor/transcription coactivator p75) and that this protein stabilizes the association of integrase with DNA (7). The interaction between integrase and LEDGF/p75 was confirmed with recombinant proteins by solution methods, such as pull-down assays (8–10), gel filtration chromatography, and analytical ultracentrifugation (2), eliminating the possibility that the interaction is weak, indirect, or mediated by DNA or another protein. The integrase-interacting domain of LEDGF/p75 was mapped independently by two groups using cell-based (11) and *in vitro* and bioinformatics (9) methods. Cherepanov *et al.* (2) report the crystal structure of this domain and the catalytic subdomain of HIV-1 integrase. Both HIV-1 and feline immunodeficiency virus PICs can be immunoprecipated with anti-LEDGF/p75 antibodies, suggesting that LEDGF/p75 is indeed a component of these PICs (12). But what is the function of this interaction? The answer to this question is not clear, but many groups are currently investigating its possible effect on integrase activity, PIC nuclear import and trafficking, and integration site target selection.

Does LEDGF/p75 play a role in the catalytic steps of breaking and joining DNA? HIV-1 integrase carries out these reaction steps *in vitro* without the assistance of other proteins, but the reactions lack the full fidelity of integration *in vivo*. In particular, many reaction products result from insertion of only one viral DNA end into one strand of target DNA rather than concerted integration of pairs of viral DNA ends. LEDGF/p75 has been reported to stimulate integration in *in vitro* assays that do not distinguish single-end integration from concerted integration (7). It would be interesting to determine whether LEDGF/p75 influences the fidelity of HIV-1 DNA integration. However, other retroviral integrases, such as RSV, do not interact with LEDGF (13), so a potential role in the catalytic steps of integration is not a universal paradigm.

Another possible role for auxiliary PIC proteins, such as LEDGF/p75, is in nuclear import and targeting. Retroviral PICs must enter the host nucleus to access their target, and, because lentiviruses like HIV-1 efficiently infect nondividing cells, these PICs must in theory traverse intact nuclear membranes. Transport of particles into and out of

Conflict of interest statement: No conflicts declared.

See companion article on page 17308 in issue 48 of volume 102.

*To whom correspondence should be addressed. E-mail: bobc@helix.nih.gov.

the nucleus is controlled by the pore complexes found in the nuclear envelope, which permit passive diffusion of particles <9 nm. The HIV-1 PIC, which is significantly larger, must therefore be actively transported into the nucleus and harbor a nuclear localization signal (NLS) (14, 15). Where is the NLS in the HIV-1 PIC? Attempts to define an NLS that is required for nuclear import of HIV-1 PICs have yielded inconclusive results. Although some individual components of the PIC, such as integrase, matrix, and Vpr, carry potential NLS signals, none of these has been shown to be essential for the nuclear import of HIV-1 PICs. LEDGF/p75 is an alternative candidate for directing nuclear import of HIV-1 PICs, although the data are far from unambiguous.

When HIV-1 integrase is expressed in the absence of other viral proteins, it resides in the nucleus (7), where it forms a complex with LEDGF/p75. The nuclear distribution pattern of integrase mutants during interphase and of the integrase mutants binding to condensed chromosomes in mitosis correlates with their colocalization with LEDGF/p75 (8). RNA interference knockdown of LEDGF/p75 leads to the diffuse redistribution of integrase in cells (8). These data suggest that LEDGF/p75 could be the factor responsible for the nuclear import of integrase. To further address this question, Maertens *et al.* (16) used deletion and missence mutagenesis to isolate the NLS in LEDGF/p75; they found a canonical simian virus 40-like NLS that transferred nuclear import activity to an otherwise diffuse β-galactosidase enzyme. A single amino acid change in this NLS was sufficient to exclude mutant LEDGF/p75 from the nucleus and to abolish nuclear import of HIV-1 integrase (16). However, Q168A, an integrase mutation that disrupts the interaction with LEDGD/p75, fails to abolish nuclear import of HIV-1 integrase, although it blocks replication of the virus at the integration step (17). Furthermore, depletion of LEDGF/p75 by small inhibitory RNA had no effect on viral replication or nuclear import of PICs, suggesting that if LEDGF/p75 is indeed involved in this process it is by a redundant mechanism (12).

The final role that PIC proteins can play in integration, target-site selection, has received considerable interest of late. Integration confers the potential for long persistence of the viral genome after infection. Whether the integrated provirus is a harmless or deadly addition to the genome depends on the site of integration and the potential for transcription. The mechanisms that regulate transcriptional activity of proviral DNA are not well understood but are likely to involve not only positional effects due to the chromatin environment at the site of integration but also epigenetic mechanisms. Recent findings that HIV-1, murine leukemia virus, and avian sarcoma virus differ in their preference for sites of integration (18–22) have led to considerable interest in the underlying mechanisms of targeting and the possible roles of cellular proteins. LEDGF/p75 possesses an additional function that may be relevant to target selection: it tethers integrase to DNA and chromatin (7, 12, 17). In fact, the IN Q168A integrase mutation that disrupts the interaction with LEDGF/p75 and blocks viral replication at the integration step prevents chromosomal tethering of integrase (17). LEDGF/p75 does not interact with integrases from Moloney murine leukemia virus or RSV (13), a feature that may or may not be related to a potential role in target site selection. It would be interesting to know whether depletion of LEDGF/p75 leads to a change in the distribution of HIV-1 integration sites.

The work of Cherepanov *et al.* (2) represents a major advance in the field of retroviral integration because of the unambiguous nature of the interaction and the unprecedented crystal structure of a complex between integrase and any other protein. Like all breakthrough results, this structure and the other experimental work present many unresolved questions. Beyond the remaining functional questions outlined above, some structural issues need to be addressed. It would be interesting to know the structural details of the interaction between LEDGF/p75 and the N-terminal domain of integrase, particularly in light of the fact that the structure of full-length integrase has yet to be determined. The interaction of integrase and LEDGF/p75 may represent the best hope for obtaining structural information on full-length integrase.

Finally, beyond the intellectually compelling aspects of this problem is the fact that integrase, as one of three virally encoded enzymes, is an important target for developing antiretroviral compounds. Integrase inhibitors that specifically block strand transfer are in advanced stages of development (23). A new frontier for integrase-targeting drugs would be ones that inhibit specific interactions in the PIC necessary for integration *in vivo*. Mutations of the LEDGF/p75 binding site on integrase have been shown to block viral replication, suggesting that modulation of this binding site by a small molecule or peptide could have serious effects on infection, regardless of the function of LEDGF/p75.

This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. C.M.B. is a Paul Sigler/Argoron Fellow of the Helen Hay Whitney Foundation.

1. Brown, P. O. (1997) in *Retroviruses*, eds. Coffin, J. M., Hughes, S. H. & Varmus, H. E. (Cold Spring Harbor Lab., Plainview, NY), pp. 161–203.
2. Cherepanov, P., Ambrosio, A. L. B., Rahman, S., Ellenberger, T. & Engelman, A. (2005) *Proc. Natl. Acad. Sci.* USA **102**, 17308–17313.
3. Turlure, F., Devroe, E., Silver, P. A. & Engelman, A. (2004) *Front. Biosci.* **9**, 3187–3208.
4. Farnet, C. M. & Haseltine, W. A. (1991) *J. Virol.* **65,** 1910–1915.
5. Bukrinsky, M. I., Sharova, N., McDonald, T. L., Pushkarskaya, T., Tarpley, W.G. & Stevenson, M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6125–6129.
6. Miller, M. D., Farnet, C. M. & Bushman, F. D. (1997) *J. Virol.* **71,** 5382–5390.
7. Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E. & Debyser, Z. (2003) *J. Biol. Chem.* **278,** 372–381.
8. Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z. & Engelborghs, Y. (2003) *J. Biol. Chem.* **278,** 33528–33539.
9. Cherepanov, P., Devroe, E., Silver, P. A. & Engelman, A. (2004) *J. Biol. Chem.* **279,** 48883–48892.
10. Cherepanov, P., Sun, Z., Rahman, S., Maertens, G., Wagner, G. & Engelman, A. (2005) *Nat. Struct. Mol. Biol.* **12,** 526–532.
11. Vanegas, M., Llano, M., Delgado, S., Thompson, D., Peretz, M. & Poeschla, E. (2005) *J. Cell Sci.* **118,** 1733–1743.
12. Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M. & Poeschla, E. M. (2004) *J. Virol.* **78,** 9524–9537.
13. Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F. & Debyser, Z. (2005) *J. Biol. Chem.* **280,** 17841–17847.
14. Fouchier, R. A. & Malim, M. H. (1999) *Adv. Virus Res.* **52,** 275–299.
15. Goff, S. P. (2001) *J. Gene Med.* **3,** 517–528.
16. Maertens, G., Cherepanov, P., Debyser, Z., Engelborghs, Y. & Engelman, A. (2004) *J. Biol. Chem.* **279,** 33421–33429.
17. Emiliani, S., Mousiner, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, Moisant, F., Ben-Slama, L., Witrouw, M., *et al.* (2005) *J. Biol. Chem.* **280,** 25517–25523.
18. Schroder, A. R. W., Shinn, P., Chen, H. M., Berry, C., Ecker, J. R. & Bushman, F. (2002) *Cell* **110**, 521–529.
19. Laufs, S., Gentner, B., Nagy, Z., Jauch, A., Benner, A., Naundorf, S., Kuehlcke, K., Schiedlmeier, B., Ho, A. D., Zeller, W. J. & Fruehauf, S. (2003) *Blood* **101,** 2191–2198.
20. Wu, X., Li, Y., Crise, B. & Burgess, S. M. (2003) *Science* **300,** 1749–1751.
21. Mitchell, R. S., Beitzel, B. F., Schroder, A. R., Shinn, P., Chen, H., Berry, C. C., Ecker, J. R. & Bushman, F. (2004) *PLoS Biol.* **2,** 1127–1137.
22. Narezkina, A., Taganov, K. D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, A. M. & Katz, R. A. (2004) *J. Virol.* **78,** 11656–11663.
23. Hazuda, D. J., Young, S. D., Guare, J. P, Anthony, N. J., Gomez, R. P., Wai, J. S., Vacca, J. P., Handt, L., Motzel, S. L., Klein, H. J., *et al.* (2004) *Science* **305,** 528–532.